心理科学进展 2023, Vol. 31, No. 4, 507–518
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.00507

·研究方法(Research Method)·

追踪研究中的内生性问题：来源与应对*

方俊燕[1] 温忠麟[2]
([1]广州体育学院休闲体育与管理学院，广州 510500)
([2]华南师范大学心理学院/心理应用研究中心，广州 510631)

摘 要 相对于横断研究，追踪研究中更有可能同时存在多种内生性问题来源。双变量追踪研究在心理学因果分析中发挥了重要的作用，然而其中的内生性问题却未得到应有的关注，这可能会影响推论的准确性。追踪研究中内生性问题的来源视乎模型而定，主要包括遗漏变量、变量选择和样本选择、解释变量的测量误差、动态面板和变量之间的相互关系。本文以代表性追踪模型 CLPM 为例，展示了内生性问题的影响，讨论了在原模型中运用工具变量来建模以应对内生性问题的可行性，目的是使心理学研究者能够关注追踪研究中的内生性问题，更好地运用追踪模型进行因果分析。

关键词 内生性问题，追踪研究，交叉滞后面板模型，工具变量
分类号 B841

1 引言[1]

在心理学等社会科学研究领域，建立和检验变量之间的因果关系许多时候依赖于观测数据，这些数据来源于真实世界中实施的观察性研究。已有不少研究注意到，基于观测数据的回归分析可能会受到内生性问题(endogeneity)的困扰。内生性问题是指在回归模型中解释变量和误差项之间存在相关关系，这违背了回归模型的基本假定，会导致有偏差的结果，进而影响对变量之间真实关系的判断(Bascle, 2008; Bollen, 2012; Maydeu-Olivares et al., 2020)。内生性问题几乎是所有基于观测数据的研究都无法忽视的议题，然而，以往有关内生性问题的讨论通常集中于横断研究(如：王宇，李海洋，2017; Bollen, 2012; Hill et al., 2020; Wooldridge, 2010)，目前仍未见有在追踪研究中系统地探讨内生性问题。

追踪研究(也称为纵向研究)是一种广受学界重视的探究变量之间因果关系的方法，尤其是当随机控制实验有违伦理或者难以实施的时候，基于重复测量数据的追踪研究在因果分析中发挥了重要的作用。随着“大数据”时代的到来，追踪研究越来越受到重视，其较少受到伦理因素的限制，外部效度较好，且成本较低可支持较长的研究周期，因此被广泛地运用于心理发展、行为治疗、临床咨询、语言认知、管理心理等各个心理学领域(如：熊猛 等，2020; Burns et al., 2019; Fang et al., 2022; Gates et al., 2020; Tong et al., 2019)。相对于横断研究，追踪研究中更难避免内生性问题。值得注意的是，横断研究中内生性问题来源较少，学者们通常识别出一种来源(即遗漏变量，如：黄斌 等，2022; 赵西亮，2017; 张晓敏 等，2022)或两种来源(即遗漏变量和双向因果，如刘立光，2021)。有些内生性问题来源是由数据分析模型的结构所造成的，例如动态面板是由模型中的自回归路径引起，这种结构常见于追踪模型，追踪研究中可以识别出两种或更多来源的内生性问题(如：李适源，刘爱玉，2022; 朱琪 等，2022; Lebenbaum et al., 2021; Lu et al., 2019)。不同的内生性问题来源可以累积(Bascle, 2008)，追踪研究中内生性问题对模型参数估计的影响有必要进行深入的分析。

收稿日期：2022-06-21
* 国家自然科学基金项目(32171091)资助。
通信作者：温忠麟, E-mail: wenzl@scnu.edu.cn

尽管追踪研究的实证运用日益增多，然而罕有研究者意识到其中内生性问题的存在，尤其在心理学领域的追踪研究中，几乎没有对内生性问题采取针对性的控制，也缺乏有关的方法学探讨。已有的追踪研究通常关注对一些表层问题(如遗漏变量)的统计控制，而对内生性问题这一更本质性、更根源性的问题尚缺乏针对性的探讨。本文关注追踪研究中的内生性问题，拟厘清不同的内生性问题来源，探讨其影响并提出应对策略。

2 追踪研究中的内生性问题

2.1 内生性问题概述

考虑一个简单的回归模型，包含一个解释变量 x 和一个结果变量 y，方程为：

$$y = \alpha + \beta x + \varepsilon \tag{1}$$

其中，α 为截距，β 为斜率，ε 为误差项。结果变量是内生的(endogenous)，变量值由模型内部决定；解释变量通常都假设是外生的(exogenous)，变量值由模型外部决定。

回归模型和结构方程模型中的一个基础假定是解释变量的外生性假定(exogeneity assumption)，即：$E(\varepsilon) = 0, \mathrm{cov}(\varepsilon, x) = 0$ (Hill et al., 2020; Wooldridge, 2010)。满足这一假定时，回归系数的最小二乘估计值是最佳线性无偏估计(best linear unbiased estimator, BLUE)。然而，实际中往往会出现解释变量和误差项有相关的情况，这违背了外生性假定，引起内生性问题，所得到的参数结果会是有偏的(Bollen, 2012; Falkenström et al., 2016; Gates et al., 2020)。存在内生性问题时的解释变量被称为内生性解释变量。内生性问题是基于观测数据的实证研究难以避免的重要问题(陈云松，范晓光，2010；王宇，李海洋，2017)。

2.2 内生性问题的来源

观测数据的回归分析中可能存在多种内生性问题的来源，这些来源也被称为回归分析的效度威胁(threats to the validity) (Maydeu-Olivares et al., 2020)。通常有以下几种来源：遗漏变量(omitted variable)、选择(selection)、解释变量的测量误差(measurement error in the predictors)、相互关系(reciprocal relation)和动态面板(dynamic panel)等(Bollen, 2012; Hill et al., 2020; Wooldridge, 2010)。虽然这些来源的外在表现形式各不相同，但是都会导致一个深层次的结果，即解释变量的内生性问题。

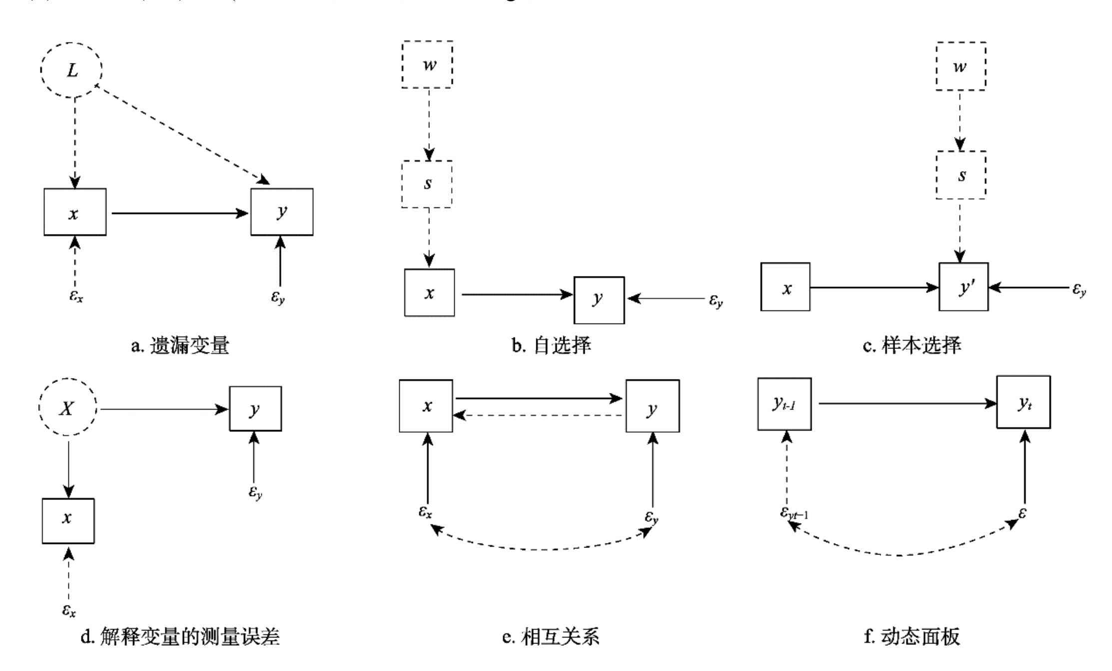


图 1 内生性问题的不同来源

注：虚线表示未被考虑进模型的变量和效应，y'表示当存在样本选择时能被测量到的那部分结果变量，L 表示遗漏变量，w 表示能够影响选择过程和作用过程的变量，s 表示选择过程。

2.2.1 遗漏变量

遗漏变量是最为常见的引起内生性问题的一个原因，是指模型设定中遗漏了某个或某些解释变量，当被遗漏的变量既影响结果变量也影响解释变量时，就可能引起内生性问题(Bollen, 2012; Hill et al., 2020)，如图 1(a)所示。遗漏变量其实就是因果关系的前置变量(温忠麟, 2017)，也称为未观测到的异质性(unobserved heterogeneity)，未观测到的偏差(unobserved bias)或未被测量的混淆变量(unmeasured confounding) (陈云松, 范晓光, 2010; 王宇, 李海洋, 2017)，当只有一个解释变量和一个结果变量时，遗漏变量也被称为第三变量(third variable)。

假设有遗漏变量 L，既影响解释变量 x 也影响结果变量 y，则有：

$$x = \alpha_x + \theta L + \varepsilon'_x \tag{2}$$

$$y = \alpha_y + \beta' x + \eta L + \varepsilon'_y \tag{3}$$

其中，α_x 和 α_y 表示截距，β' 表示 y 对 x 的回归系数，θ 和 η 分别表示 x 和 y 对 L 的回归系数，ε'_x 和 ε'_y 表示误差项。由(2)解出 L 并代入(3)可得：

$$L = (x - \alpha_x - \varepsilon'_x) / \theta \tag{4}$$

$$y = \alpha_y + \beta' x + \eta(x - \alpha_x - \varepsilon'_x) / \theta + \varepsilon_y = \alpha_y - \eta\alpha_x / \theta + (\beta' + \eta / \theta)x + (-\eta\varepsilon'_x / \theta + \varepsilon_y) \tag{5}$$

可以看出，复合误差项($-\frac{\eta}{\theta}\varepsilon'_x + \varepsilon_y$)中的 ε'_x 与 x 有相关，所以复合误差项与解释变量 x 有相关，即存在内生性问题。假设变量之间的真实关系如公式(3)所示，其中 x 对 y 的预测效应是 β'，研究者如果忽视内生性问题，不考虑遗漏变量 L 的影响，直接对方程(1)进行估计，所得到的回归系数($\beta' + \frac{\eta}{\theta}$)就会是有偏的。

遗漏变量所引起的内生性问题在心理学领域十分常见，例如，Lu 等(2019)在分析社会信任和幸福感之间的关系时，考虑到很可能会有同时影响社会信任和幸福感的变量未被测量或观测到，进而引起内生性问题，因此使用了工具变量方法来降低内生性问题的影响。

2.2.2 选择

一些文献在阐述内生性问题来源时还会提到选择，如自选择(self-selection/selection of treatment)和样本选择(sample selection/selection into sample) (Hill et al., 2020)。

自选择(Shaver, 1998)是指解释变量不是随机的，而是选择的结果(王宇, 李海洋, 2017; Hill et al., 2020)。如果将社会现象拆分为两个过程，一个过程是解释变量(x)影响结果变量(y)的过程，另一个是选择解释变量的过程(selection process, s)，若存在一些能够同时影响这两个过程的变量(w)未被观测到，作为遗漏变量而进入两个过程的误差项，使得两个误差项相互关联，导致第一个过程中的 x 和误差项有相关，内生性问题出现，如图 1(b)所示。有学者指出，在研究社区环境对个体收入的影响时难以避免自选择所带来的偏差，例如，低学历个体(w)往往只能选择搬入低档社区(x)，而低学历(w)和较低的收入(y)有关联，低学历(w)这一变量同时与解释变量和结果变量有关，如被遗漏就可能引起内生性问题(解垩, 宋颜群, 2021)。

样本选择(Heckman, 1976, 1979)是指样本不是随机的，观测仅仅局限于某个有限的非随机样本中，这可能由数据收集程序造成，也可能由所研究的某种社会现象本身所固有的特质引起。如图 1(c)所示，在考虑 x 对 y 的影响时，结果变量 y 的观测值受到选择过程(s)的影响，而 s 受到一些其他因素 w 的影响，未被观测到的因素 w 会进入方程误差项 ε_y，而 x 是 w 的子集(即 w 中至少包含一个变量不属于 x)，因此 x 和 ε_y 有相关，出现内生性问题(Hill et al., 2020)。例如，在研究农村教育收益时会关注教育年限对年收入的影响，然而受教育程度高的农村居民往往会突破户籍限制，导致通常的农村调查样本是一个高度选择性样本，往往仅包含潜在收入较低或教育水平较低者，如果使用这一部分样本进行估计，将会低估教育对农村居民收入的作用(赵西亮, 2017)。

自选择和样本选择被统称为选择。自选择偏差可以看作是混淆了感兴趣的参数和个体选择特定解释变量的概率函数的参数，样本选择可以看作是混淆了感兴趣的参数和决定个体是否进入非随机样本的概率函数的参数(Heckman, 1979)。当解释变量对结果变量的作用过程以及选择解释变量或选择被试的过程同时受到一个或多个遗漏变量影响时，这两个过程的误差项就会存在相关，导致第一个过程中的解释变量和误差项相关，出现内生性问题。从这个角度看，选择可以看作是特殊的遗漏变量(陈云松, 范晓光, 2010; Heckman,

1976)。

2.2.3 解释变量的测量误差

如果解释变量存在测量误差，按照因子分析的思路，解释变量就会受到一个潜变量(因子)的影响，因而原本的解释变量成了内生性变量。换一个角度看，这相当于模型中遗漏了一个潜变量，该潜变量既会影响解释变量也会影响结果变量(Bollen, 2012)，如图 1(d)所示。然而，测量误差造成的问题，实质上是信度高低的问题(温忠麟 等, 2022)。通过多个题目指标合成总分进行分析，合成信度通常都比较高，因而使用合成分数进行显变量建模时并不总要考虑由测量误差引起的内生性问题。当测量信度低时，可能存在测量误差引起的内生性问题，此时可以通过工具变量方法处理(Hill et al., 2020)。

2.2.4 相互关系

相互关系是指解释变量与结果变量相互预测的情况，也称为反馈因果(feedback causation)、联立性/双向因果(simultaneity) 或反向因果(reverse causality) (王宇, 李海洋, 2017; Bollen, 2012; Hill et al., 2020)。值得注意的是，有学者认为基于观测数据建立的时序关系本质上还是相关性的(Usami et al., 2019)，因果推论往往还是需要有理论、文献或经验常识的支持(温忠麟, 2017)，理论研究者们在使用观测数据探究变量间相互预测的关系时更多使用相互关系(reciprocal relation)而非因果关系(causality) (如 Usami et al., 2019; Wiedermann & von Eye, 2020)，因此本文也采用相互关系这一表述。

如图 1(e)所示，当两个变量之间存在相互关系时，x 影响 y 的方程中的误差项就会和 y 影响 x 的方程中的误差项存在关联，进而导致第一个方程中的解释变量和误差项有相关，出现内生性问题。变量之间的相互关系在心理学领域中十分常见，例如，Lebenbaum 等(2021)在研究社会资本与心理健康之间的影响效应时考虑了二者之间可能存在相互关系，这很可能引起内生性问题，因此他们采用工具变量方法来处理内生性问题并获得可靠的结果。

2.2.5 动态面板

动态面板(也称为自回归，auto-regression)是指使用结果变量的先前取值作为解释变量(Patel & Cooper, 2014)，动态面板所引起的误差也被称为自回归误差(Bollen, 2012)。在追踪研究中，变量的当前水平往往受到其本身上一时间点水平的影响，因此方程中往往会将结果变量的先前水平作为一个解释变量纳入，以控制一个潜在的遗漏变量的影响(Cole & Maxwell, 2003)，但这同时也可能引起内生性问题。如图 1(f)所示，由于结果变量的当前误差项 ε_{yt} 与上一时间点的误差项 ε_{yt-1} 有可能是相关的，而上一时间点的误差项 ε_{yt-1} 与上一时间点的结果变量 y_{t-1} 相关，这样，以 y_t 为结果变量的方程中误差项 ε_{yt} 有可能与其中一个解释变量 y_{t-1} 有相关，出现内生性问题(Bollen, 2012)。

可以看出，内生性问题可能由多种来源引起，下面将针对追踪研究情境进行详细论述。

2.3 追踪研究中的内生性问题

目前存在多种追踪模型，例如潜在曲线模型(Latent Curve Model, LCM)，交叉滞后面板模型(Cross-Lagged Panel Model, CLPM)，随机截距交叉滞后面板模型(Random Intercept CLPM, RI-CLPM)和潜在变化分数模型(Latent Change Score Model, LCS)等。其中，CLPM、RI-CLPM 和 LCS 都属于交叉滞后追踪模型，这一类模型是公认的探究变量之间动态关系的强有力的方法，也是目前运用最广泛的追踪研究技术之一(如：方俊燕 等, 印刷中; 熊猛 等, 2020; Burns et al., 2019; Fang et al., 2022; Tong et al., 2019)。

相比于其它的追踪模型，交叉滞后追踪模型尤其适合分析生命历程中两个变量之间的动态相互关系，这是心理学研究领域中广受关注的主题(Hamaker et al., 2015; Usami et al., 2019)，同时，交叉滞后追踪模型的结构也比其他追踪模型要复杂(见图 2)，可能同时面临多种不同的内生性问题来源。这里将针对交叉滞后追踪模型的基础形式 CLPM 来分析内生性问题，所得到的结果能够涵盖其它结构较为简单的追踪模型，包括并不限于 LCM。

在 CLPM 中，对于变量 x 和 y，被试 i (i = 1, …, N)在时间点 t (t = 1, …, T)的观测值为 x_{it} 和 y_{it}：

$$x_{it} = \mu_{xt} + x_{it}^{*} \quad (6)$$

$$y_{it} = \mu_{yt} + y_{it}^{*} \quad (7)$$

其中，μ_{xt}、μ_{yt} 表示组均值。首先对观测变量进行中心化处理，提取出时间特定的偏移量 x_{it}^{*}、y_{it}^{*}，以表示每个被试在不同变量上随时间的变动，

$$x_{it}^{*} = x_{it} - \mu_{xt} \quad (8)$$

$$y_{it}^{*} = y_{it} - \mu_{yt} \quad (9)$$

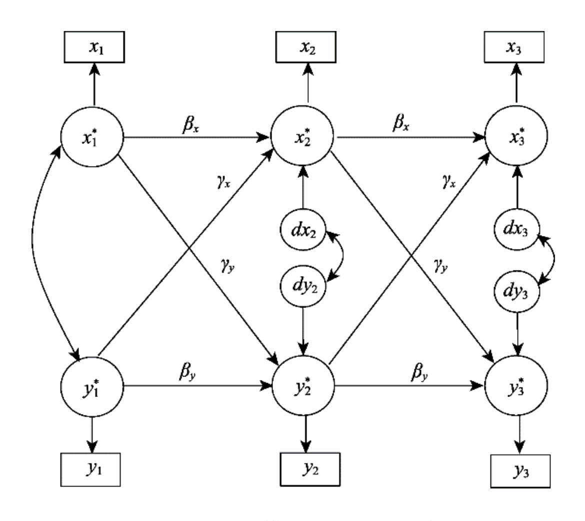


图 2　CLPM 模型图(3 个时间点)

注：两个变量为 x 和 y，下标表示测量时间点，β 为自回归效应，γ 为交叉滞后效应，d 为残差。

然后构建交叉滞后方程：

$$x_{it}^* = \beta_{xt} x_{i(t-1)}^* + \gamma_{xt} y_{i(t-1)}^* + d_{xit} \tag{10}$$

$$y_{it}^* = \beta_{yt} y_{i(t-1)}^* + \gamma_{yt} x_{i(t-1)}^* + d_{yit} \tag{11}$$

其中，β_{xt}、β_{yt} 为自回归系数，反映了一个变量的先前水平对该变量当前水平的作用，γ_{xt}、γ_{yt} 为交叉滞后系数，体现了一个变量的先前水平对另一变量当前水平的作用，d_{xit}、d_{yit} 为误差项。

首先，CLPM 难以避免遗漏变量所带来的内生性问题。假设有遗漏变量 L_{it}，既影响 x_{it}^* 也影响 y_{it}^*。如果遗漏变量因为难以施测或未被考虑到而无法获得其观测值，可能会被包含在误差项里(Bollen, 2012)，则公式(10)和(11)可改写为：

$$x_{it}^* = \beta'_{xt} x_{i(t-1)}^* + \gamma'_{xt} y_{i(t-1)}^* + (d'_{xit} + \theta_{xt} L_{it}) \tag{12}$$

$$y_{it}^* = \beta'_{yt} y_{i(t-1)}^* + \gamma'_{yt} x_{i(t-1)}^* + (d'_{yit} + \eta_{yt} L_{it}) \tag{13}$$

以 y_{it}^* 为结果变量的方程为例，由于遗漏变量 L_{it} 无法作为解释变量进入方程，因而被包含进误差项，使得原来的误差项 d_{yit} 变成复合误差项($d'_{yit} + \eta_{yt} L_{it}$)，已知 L_{it} 与 x_{it}^* 有相关，x_{it}^* 和其本身的历史变量 $x_{i(t-1)}^*$ 有相关，则复合误差项($d'_{yit} + \eta_{yt} L_{it}$)可能与 $x_{i(t-1)}^*$ 有相关，产生内生性问题。在运用 CLPM 的研究中也有可能会面临由样本选择或自选择所引起的内生性问题，考虑到选择可以看作是遗漏变量的特殊情形，不再单独论述。

其次，CLPM 通常都是基于合成分数来建模的，当量表信度较高的时候可以不考虑由测量误差所引起的内生性问题。

再次，CLPM 中自回归效应的存在可能会导致由动态面板引起的内生性问题。考虑 y_{it}^* 为结果变量的回归方程，由公式(11)可以看出，y_{it}^* 的其中一个解释变量是其本身在(t−1)时刻的历史变量 $y_{i(t-1)}^*$，已知 $y_{i(t-1)}^*$ 和 $d_{yi(t-1)}$ 有相关，由于回归方程的误差是一直存在的，得到 $d_{yi(t-1)}$ 和 d_{yit} 可能有相关(Bollen, 2012)，则 $y_{i(t-1)}^*$ 和 d_{yit} 很可能有相关，产生内生性问题。

最后，CLPM 对变量之间的相互关系进行建模，可能面临由相互关系引起的内生性问题。考虑 y_{it}^* 为结果变量的方程，由于 d_{yit} 和 d_{xit} 有相关，而 d_{xit} 和 x_{it}^* 有相关，x_{it}^* 和其本身的历史变量 $x_{i(t-1)}^*$ 有相关，则 d_{yit} 和 $x_{i(t-1)}^*$ 可能有相关；同理，在 x_{it}^* 为结果变量的方程中，d_{xit} 和 $y_{i(t-1)}^*$ 可能有相关，上述两种情形任一种成立都有可能引起内生性问题。当两种情形都成立时，变量之间存在相互关系；仅有一种情形成立时，变量之间是单向关系，可看作相互关系的特例。

综上，在以 CLPM 为代表的追踪模型中可能存在以下几种内生性问题来源：遗漏变量，动态面板和相互关系。其中，遗漏变量是横断研究和追踪研究中都常见的，几乎所有的追踪模型(如 LCM、CLPM 等)都可能碰到遗漏变量引起的内生性问题；动态面板则是由追踪模型的结构所决定的，在包含自回归路径的追踪模型(如 RI-CLPM、LCS)中都会存在动态面板引起的内生性问题；在变量之间存在相互作用的背景下开展的研究中往往会存在相互关系引起的内生性问题。在实际研究情境中，可以采用不同的追踪模型，再结合理论背景和模型结构来识别不同的内生性问题来源。

2.4　内生性问题的影响

内生性问题会导致常用的回归分析方法得到有偏差、不一致的结果，进而影响对变量之间真实关系的判断(Bollen, 2012)。重要的是，内生性问题所导致的偏差无法被预测，而且参数被高估的可能性和被低估的可能性是一样的(Bollen, 2012; Hill et al., 2020)。在某些情况下，内生性问题会导致研究者错误地得出原本并不存在的效应，而有些情况下会令研究者无法发现原本存在的效应(Certo et al., 2016)。例如，Lu 等(2019)的横截面研究中发现在修正内生性问题之后，社会信任对幸福感的预测效应由原本的 0.198 (p < 0.01)提升至

0.505 (p < 0.001)。

已有的研究往往是围绕横截面情境中的内生性问题展开(如：陈云松, 范晓光, 2010; 王宇, 李海洋, 2017; 赵西亮, 2017; Bollen, 2012; Hill et al., 2020), 目前仍未见到在追踪情境下对内生性问题的系统探讨。追踪研究中可能面临哪些内生性问题来源？内生性问题对模型参数估计的影响是怎样的？此时何种应对方法能有效应对内生性问题？这都是有待探究的问题。

3 内生性问题的应对方法

3.1 实验或准实验控制

内生性问题可以通过研究设计、实验或准实验等方式来进行校正(Hill et al., 2020)。

针对遗漏变量引起的内生性问题, 可以识别出遗漏变量并将其纳入模型中控制, 也可以在模型中纳入协变量来代表可能的遗漏变量(Usami et al., 2019), 在实证研究中, 学者们往往对人口学变量进行控制(如：熊猛 等, 2020)。然而, 要确定合适的协变量并获得其有效观测值, 对追踪研究的设计和实施有较高的要求, 如未能纳入对研究变量有实质性影响的协变量, 可能无法从根本上解决遗漏变量所引起的内生性问题。对于选择引起的内生性问题, 可以采用准实验的方式进行控制, 例如为了研究社区邻里效应, 有研究者在城市里向随机选择的家庭提供住房券, 以获得合理的实验组和对照组(Aliprantis & Richter, 2020)。

当实验控制方法由于成本等原因无法实现或较难实施时, 可以采用统计方法进行修正。

3.2 统计控制

目前存在一系列能够对内生性问题进行修正的统计方法, 例如工具变量方法(instrumental variables method, IV method), Heckman 两阶段模型, 双重差分分析(difference in differences analysis), 倾向性得分匹配(propensity score matching)和固定效应模型(fixed effect model)等(王宇, 李海洋, 2017)。IV 方法是通过 IV 来提取出解释变量的外生部分进行估计; Heckman 两阶段模型主要用于修正选择偏差, 第一阶段为概率模型, 估计自选择偏差变量发生的可能性, 得到比率加入到第二阶段模型中; 双重差分分析是使用两次差分来获得样本在实验处理前后的差异; 倾向性匹配得分则是采用倾向得分(发生概率)进行匹配以重新建立自然实验的条件(王宇, 李海洋, 2017)。其中, IV 方法的适用性较强, 能应对多种内生性问题来源, 是实证中最常用的技术(Bollen, 2012; Hill et al., 2020; Maydeu-Olivares et al., 2020)。

IV 是解释变量的解释变量。IV 与通常所说的控制变量不是一回事, 控制变量是指除了解释变量以外的所有影响实验结果的变量, 通常会对实验的遗漏变量进行控制, IV 则是纳入估计模型中的辅助性变量, 能处理多种原因所引起的内生性问题, 包括且不限于遗漏变量。合格的 IV 需要满足两个条件：①关联性, IV 与内生性解释变量有强相关, 即 $\mathrm{cov}(x,z) \neq 0$, 其中 cov 表示协方差(covariance); ②外生性, IV 与误差项不相关, 即 $\mathrm{cov}(\varepsilon,z) = 0$ (Bollen, 2012; Hill et al., 2020; Maydeu-Olivares et al., 2020)。IV 模型(在原模型中纳入 IV 之后的模型)可以运用两阶段最小二乘回归法(two-stage least-squares regression, 2SLS)来估计：第一阶段, 将原本的内生性解释变量作为因变量, 建立解释变量对 IV 的回归, 获得解释变量的拟合值; 第二阶段, 建立结果变量对解释变量拟合值的回归, 此时使用解释变量的外生部分(即与误差项不相关的部分)来估计, 实现了对内生性问题的控制, 可获得反映变量间真实关系的系数(Bollen, 2012; Hill et al., 2020)。IV 模型也可以采用极大似然估计(maximum likelihood, ML)和广义矩估计(generalized method of moments, GMM)等方法估计。Maydeu-Olivares 等(2020)发现 2SLS 仅仅是 ML 的一个多阶段形式, 在没有模型误设的情况下所得到的参数结果到小数点的后两位都是一致的。特别地, ML 是心理学有关领域的学者们更为熟悉的一种参数估计方法。

IV 的质量评估通常包含弱 IV 检验(weak IV check)和过识别检验(overidentification test)。弱 IV 检验是针对关联性前提, 当模型中包含一个 IV 时, 可以对解释变量和 IV 之间的相关系数或回归系数进行显著性检验; 当模型中包含多个 IV 时, 存在多种检验方法, 可以针对每一个内生性解释变量计算一个偏 R^2 来判断 IV 是否满足关联性(Bollen, 2012; Shea, 1997), 也可以通过计算 IV 模型的 F 统计量或比较不同 IV 模型中 T_{CD} 统计量的特征值来检验(Maydeu-Olivares et al., 2020), 同时, Maydeu-Olivares 等(2020)建议可以使用嵌套模型比较的方法, 即通过似然比检验来比较 IV 模

型和零模型(即 IV 对解释变量的回归系数设置为零), 进而判断 IV 是否满足关联性要求。过识别检验的原假设是所有 IV 都和方程误差项不相关, 适用于 IV 个数多于解释变量个数时(Bollen, 2012), 一般通过 Sargan 检验进行, 即构建解释变量对所有 IV 的回归, 使用回归误差项和 IV 的取值构造一个服从卡方分布的统计量 T_S 并进行判断, 此外, Maydeu-Olivares 等(2020)指出, 在 SEM 框架下运用 ML 方法时, 对于 IV 和误差项不相关这一假设, 可以通过模型拟合的卡方检验进行验证, 如果模型拟合得较好就不会拒绝该假设。

关于实证中应当采用的 IV 的个数, 目前尚未有一致的结论。有学者认为 IV 的个数至少要和内生性解释变量的个数一样多(Bollen, 2012; Falkenström et al., 2016), 也有学者认为应当比内生性解释变量多一个(Maydeu-Olivares et al., 2020)。

在社会科学领域, IV 的常见类型有两种, 分别是辅助的 IV (auxiliary IV, AIV)和模型隐含的 IV (model implied IV, MIIV), AIV 是从外部引入模型中的, MIIV 则是来源于模型内部(Bollen, 2018; Hill et al., 2020; Semadeni et al., 2014)。有效的 AIV 不易找到, 构建和选择 AIV 都需要十分严密的逻辑, 否则其有效性容易受到质疑(陈云松, 2012)。相比之下, MIIV 具有一些不可替代的优点, 其来源于模型内部, 稳定性较高, 且模型的结构表明了模型中哪些变量满足成为 IV 的条件, 可以不用费力地辨别 MIIV。学者们提出, 解释变量的历史取值(简称历史变量)往往是适当的 IV, 历史变量和当前的解释变量有相关, 而几乎不可能与当前的误差项有相关, 即其满足 IV 的要求(Gates et al., 2020; Hill et al., 2020), 追踪数据的重复测量结构为使用历史变量作为 IV 提供了机会(Streeter et al., 2017), 此时历史变量属于 MIIV。

可以看出, 对于追踪研究中的内生性问题, IV 方法, 尤其是 MIIV 是一种十分有前景的处理方法, 但目前仍未见有将 MIIV 运用于追踪研究中的尝试, 因此, 有必要探究在追踪模型中纳入 MIIV 的建模方式, 分析其可行性和有效性, 为实证研究者提供切实可行的指导。

4　实证示例

下面将基于一个实证案例, 来展示内生性问题对追踪模型参数估计的影响, 并分析在模型中运用 MIIV 的效能。此处选择广受关注的代表性追踪模型 CLPM 进行演示。

数据来源于中国健康与营养调查(China Health and Nutrition Survey, CHNS)的开放数据库, 拟探究主观幸福感和身体质量指数(Body Mass Index, BMI)之间的纵向关系, 主观幸福感由一道问卷题目测量, 其要求被试按照 1(非常好)到 5(非常差)的等级评价自己当前的生活, 反向编码之后得分越高表明幸福感越高; BMI 由被试自我报告或他人代答的身高和体重数据来计算：BMI = 体重(千克)/身高(米)的平方。选取三个时间点(T1 = 2009, T2 = 2011, T3 = 2015), 删除异常值并保留那些在三个时间点都有记录的被试, 最终获得来自 992 名被试的有效数据, 在初始时间点上被试平均年龄为 51 岁(范围为 18~88 岁), 57%为女性。

变量在不同时间点上的描述性统计和相关关系见表 1。

首先, 需要识别研究中的内生性问题来源。其一, 遗漏变量, 可能会有变量同时影响主观幸福感和 BMI, 例如人口学变量、家庭社会经济地位、文化背景等(Fang et al., 2022; Kazuma, 2021), 在一次分析中无法穷尽所有可能的变量进行控制, 因此无法肯定地排除遗漏变量的存在; 其二, 动

表 1　变量的描述性统计和相关关系

变量	*M* (*SD*)	幸福感(T1)	幸福感(T2)	幸福感(T3)	BMI (T1)	BMI (T2)	BMI (T3)
幸福感(T1)	3.51 (0.83)	1					
幸福感(T2)	3.63 (0.87)	0.34^{**}	1				
幸福感(T3)	3.62 (0.83)	0.23^{**}	0.30^{**}	1			
BMI (T1)	23.53 (6.10)	0.10^{**}	0.17^{**}	0.13^{**}	1		
BMI (T2)	23.80 (5.72)	0.10^{**}	0.18^{**}	0.14^{**}	0.95^{**}	1	
BMI (T3)	24.02 (5.58)	0.10^{**}	0.15^{**}	0.11^{**}	0.89^{**}	0.89^{**}	1

注：*M*, 均值; *SD*, 标准差; **, $p < 0.01$。

态面板，此处使用交叉滞后追踪模型来分析两个变量之间的纵向关系，由于模型中默认包含自回归效应，可能面临由动态面板引起的内生性问题；其三，相互关系，有研究指出主观幸福感和 BMI 之间存在互相预测的关系(如 Fang et al., 2022)，因此本研究可能面临由相互关系引起的内生性问题。可以看出，本研究可能会面临遗漏变量、动态面板和相互关系这三种内生性问题来源。

其次，确认分析模型，将使用 CLPM 进行分析，同时，考虑到在回归分析中可能会有一些影响结果变量的人口学变量同时也影响解释变量，需要进行控制(温忠麟, 2017)，这里使用性别和年龄作为协变量纳入 CLPM，得到的模型记为 CLPM-cov。此外，我们使用 IV 方法来处理内生性问题，考虑到有效的 AIV 不易找到，且 MIIV 相比于 AIV 具有一些不可替代的优点，这里通过在 CLPM 中纳入 MIIV 来运用 IV 方法，获得的新模型记为 MIIV-CLPM。采用稳定性假设，不同时间点上同一类型的路径系数设置为不随时间变化，因此仅针对最后一个时间点的方程运用 MIIV，在 T3 时间点的两个方程中结果变量分别为 x_{i3}^* 和 y_{i3}^*，解释变量都是 x_{i2}^* 和 y_{i2}^* (见公式 10 和 11)。注意到每一个方程中的解释变量都是两个，因此 MIIV 至少也需要两个(Bollen, 2012)。前文提到，历史变量往往是适当的 IV (Gates et al., 2020)，也有通过一系列模拟研究表明在 CLPM 中使用历史变量作为 MIIV 来建模能够有效应对内生性问题，获得参数的无偏估计(Fang et al., under review)，此处将使用历史变量作为 MIIV。在 T3 时间点的方程中解释变量的历史变量是 x_{i1}^* 和 y_{i1}^*，考虑到 IV 模型往往都是基于观测变量来建模(Bollen, 2018; Maydeu-Olivares et al., 2020)，因此选用两个历史变量的观测值(即 y_1 和 x_1)作为 MIIV，它们与解释变量(x_{i2}^* 和 y_{i2}^*)有相关，且与误差项(d_{xi3} 和 d_{yi3})无相关，满足 MIIV 的要求。具体而言，参考 Maydeu-Olivares 等(2020)的 IV 模型构建方法，在 T3 时间点的方程中构建解释变量(x_{i2}^* 和 y_{i2}^*)对 MIIV(y_1 和 x_1)的回归路径(记为 z)，MIIV 是满足外生性要求的，这样获得的回归预测值就是解释变量的外生部分，将其作为新的解释变量来进行估计，同时设置解释变量误差项和结果变量误差项间的相关，就能避免内生性问题，满足回归分析的要求(见图 3)。所有分析通过 R 软件中的 lavvan 包(Rosseel, 2012)实现(见网络版附录)。

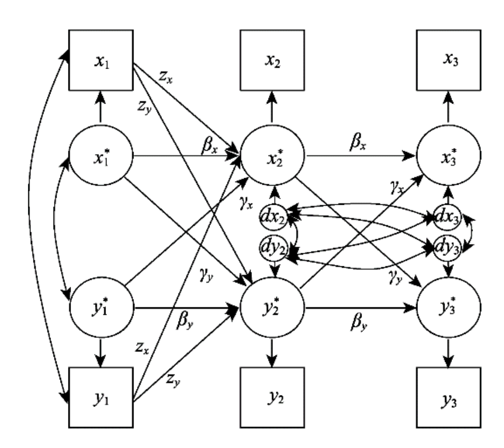


图 3 在 T3 时间点运用 MIIV

注：z 为解释变量(即 x_{i2}^* 和 y_{i2}^*)对 MIIV(即 y_1 和 x_1)的回归系数。

不同模型的拟合结果见表 2。当拟合指数满足以下条件：RMSEA 小于 0.08, SRMR 小于 0.08, CFI 大于 0.90, TLI 大于 0.90，可以认为模型能够较好地拟合数据(温忠麟 等, 2004)。

然后，对估计模型的结果进行比较，选出较适当的模型作为最终模型。

拟合表现方面，MIIV-CLPM 的拟合表现显著地优于 CLPM ($\Delta\chi^2 = 130.159$, $\Delta df = 7$, $p < 0.05$; ΔRMSEA > 0.05)和 CLPM-cov ($\Delta\chi^2 = 152.545$, $\Delta df = 15$, $p < 0.05$; ΔRMSEA > 0.05), CLPM-cov 的拟合表现好于原模型，但差于 MIIV-CLPM，这表明本例中纳入协变量能够一定程度上改善模型拟合，然而相比之下，使用 IV 方法来建模仍然是更优秀的方案。使用历史变量作为 MIIV 能够满足外生性前提，同时 MIIV-CLPM 良好的拟合表现也验证了所用的 MIIV 满足外生性前提(Maydeu-Olivares et al., 2020)。此外，MIIV 对解释变量的预测作用都是显著的($ps < 0.01$)。如果将 MIIV-CLPM 中 MIIV 对解释变量的预测效应固定为零，可以获得一个新模型 $\text{MIIV-CLPM}_{\text{restrict}}$，MIIV-CLPM 的拟合表现优于新的限制模型($\chi^2 = 119.018$, $df = 4$, CFI = 0.972, RMSEA = 0.170)，说明 MIIV 与解释变量有实质性关联(Maydeu-Olivares et al., 2020)，即 MIIV 能够满足关联性前提。

参数结果方面，CLPM 中变量的自回归效应都是显著的($p < 0.001$)，这说明主观幸福感和 BMI 都具有较高的跨时间稳定性；在控制了变量

表 2　不同模型的拟合表现和参数估计结果

系数/拟合指标	CLPM			CLPM-cov			MIIV-CLPM		
	Est.	*SE*	*p*	Est.	*SE*	*p*	Est.	*SE*	*p*
β_{x1}	0.71	0.09	<0.001	0.73	0.09	<0.001	0.59	0.09	<0.001
β_{x2}	0.70	0.09	<0.001	0.72	0.09	<0.001	0.69	0.09	<0.001
β_{y1}	0.97	0.01	<0.001	0.97	0.01	<0.001	1.00	0.02	<0.001
β_{y2}	0.97	0.01	<0.001	0.97	0.01	<0.001	0.94	0.02	<0.001
γ_{x1}	0.00	0.12	0.898	0.00	0.12	0.810	0.03	0.24	0.291
γ_{x2}	0.00	0.12	0.898	0.00	0.12	0.810	0.04	0.24	0.291
γ_{y1}	0.10	0.00	0.005	0.10	0.00	0.007	0.02	0.01	0.661
γ_{y2}	0.09	0.00	0.005	0.09	0.00	0.007	0.02	0.01	0.661
θ_x	—	—	—	–0.03	0.05	0.523	—	—	—
θ_y	—	—	—	0.10	0.39	0.003	—	—	—
η_x	—	—	—	–0.03	0.00	0.548	—	—	—
η_y	—	—	—	0.05	0.01	0.132	—	—	—
z_{x1}	—	—	—	—	—	—	0.03	0.01	0.001
z_{x2}	—	—	—	—	—	—	0.21	0.01	0.001
z_{y1}	—	—	—	—	—	—	0.05	0.03	<0.001
z_{y2}	—	—	—	—	—	—	0.37	0.03	<0.001
χ^2		130.371			152.757			0.212	
df		9			17			2	
CFI		0.971			0.967			1.000	
RMSEA		0.117			0.090			0.000	

注：*x*, 主观幸福感; *y*, BMI;
Est., 标准化路径系数, 稳定性假设仅针对非标准化结果, 而标准化结果中不同时间点的系数可能不同;
SE, 标准误; *p*, 显著性水平;
β_{xt} , *x* 的自回归效应; β_{yt} , *y* 的自回归效应; γ_{xt} , *y* 对 *x* 的滞后影响效应; γ_{yt} , *x* 对 *y* 的滞后影响效应;
θ_x , θ_y , 性别对 *x* 或 *y* 的预测作用; η_x , η_y , 年龄对 *x* 或 *y* 的预测作用;
z_{x1} , 第 1 个 MIIV(x_1)对 *x* 的预测作用; z_{x2} , 第 2 个 MIIV(y_1)对 *x* 的预测作用;
z_{y1} , 第 1 个 MIIV(x_1)对 y 的预测作用; z_{y2} , 第 2 个 MIIV(y_1)对 *y* 的预测作用。

本身上一时间点的影响之后, 两个变量之间存在单向的负向交叉滞后效应(γ_{yt} = 0.09~0.10, $p < 0.01$), 表明幸福感能够负向影响下一时间点的 BMI。使用了人口学变量作为协变量的 CLPM-cov 所得到的结果与原模型基本一致, 但是其中一个协变量(年龄)的作用并不显著, 在实证中要找到对研究变量有实质性影响的协变量是不容易的, 能获得协变量的有效测量来建模更是对研究设计提出了较高的要求, 通常一些潜在的遗漏变量的完整测量是很难获得的, 例如家庭净收入数据在此处使用的追踪调查中缺失高达 66% (T1), 67% (T2, T3)。在 MIIV-CLPM 中, 变量的自回归效应也都是显著的, 但是交叉滞后效应都不显著, 说明幸福感和 BMI 之间并不存在相互预测关系。

可以看出, MIIV-CLPM 的统计表现优秀, 显著好于其他模型。此外, CLPM 忽视了内生性问题, CLPM-cov 纳入了一些协变量但无法确切地排除其他遗漏变量的可能, 且还可能面临动态面板和相互关系所引起的内生性问题, 而 MIIV-CLPM 通过运用 IV 方法从根本上对内生性问题进行了处理, 其所得到的估计结果是更加可信的。因此, 此处选择 MIIV-CLPM 作为最终模型, 结果表明主观幸福感和 BMI 之间不存在影响关系。

综上, 在追踪研究中可能会同时存在多种内

生性问题来源，不同的来源视乎模型而定，在CLPM中可能存在遗漏变量、动态面板和相互关系，其他包含交叉滞后结构的模型中(如RI-CLPM和LCS等)同样可能面临这些内生性问题来源，在不包含自回归路径的追踪模型(如LCM)中不存在动态面板这一内生性问题来源，可能存在遗漏变量，是否存在相互关系则需要结合所研究变量的理论背景和具体模型结构来作判断。在分析变量间的纵向预测关系时，内生性问题的存在可能会导致估计结果出现偏差。在本例中，CLPM由于内生性问题的影响得到了原本并不存在的效应，当采用不同的研究变量或其他的追踪模型进行分析时，这种情况是否仍会出现还有待未来进一步的拓展研究和验证。此外，本例中纳入协变量后模型的拟合表现差于MIIV模型，这在不同研究情境下也是需要具体分析的。未来研究也可以尝试探讨在不同的变量类型下IV方法的表现是否有差异。重要的是，研究者应当意识到内生性问题的存在及其对估计结果的影响，有意识地采取应对措施。本文的示例表明，在追踪模型的原模型中运用MIIV来建模是一种能够应对内生性问题的可行途径。

5 结论与建议

基于观测数据的追踪研究被广泛地用于变量之间因果关系的分析，然而其中的内生性问题却未得到应有的关注，这可能会导致参数估计偏差，进而影响因果推断的准确性。本文关注追踪研究中的内生性问题，厘清了不同的内生性问题来源，以近年来广受关注的追踪模型CLPM为例，展示了内生性问题的影响，率先将IV方法引入CLPM中应对内生性问题，证实了其可行性和有效性，为IV方法在其他追踪模型中的推广应用奠定了基础。

首先，追踪研究中可能存在多种不同的内生性问题来源，视乎模型而定，CLPM中可能面临的来源主要包括遗漏变量、动态面板和相互关系；

其次，内生性问题会导致追踪分析的参数估计出现偏差，例如本文的实例中内生性问题令CLPM在估计交叉滞后系数时得到原本并不存在的效应；

最后，IV方法是应对追踪研究中内生性问题的可行方案，尤其是MIIV能够充分利用已有的数据来得到可靠的参数估计结果。

综上，实证研究者应当意识到追踪研究中内生性问题的存在，并采取适当的应对措施，建议在追踪模型中使用历史变量作为MIIV来处理内生性问题，进而揭示变量间的真实关系。

参考文献

陈云松. (2012). 逻辑、想象和诠释：工具变量在社会科学因果推断中的应用. *社会学研究*, *27*(6), 192–216.

陈云松，范晓光. (2010). 社会学定量分析中的内生性问题 测估社会互动的因果效应研究综述. *社会*, *30*(4), 91–117.

方俊燕，温忠麟，黄国敏. (印刷中). 纵向关系的探究：基于交叉滞后结构的追踪模型. *心理科学*.

黄斌，何沛芸，朱宇，魏易. (2022). 基于父母教育背景的中国家庭校外教育支出分化——兼论家庭需求视角下“双减”政策实施的优化. *中国教育学刊*, (4), 19–28.

解垩，宋颜群. (2021). 收入的社区邻里效应研究. *经济管理*, *43*(6),190–208.

李适源，刘爱玉. (2022).“忧郁的孩子们”：课外补习会带来负向情绪吗？基于中国教育追踪调查(CEPS)两期数据的因果推断. *社会*, *42*(2), 60–93.

刘立光. (2021). 精神健康能否提升老年人主观幸福感？——兼论社会互动的中介效应. *科学经济社会*, *39*(4), 56–68.

王宇，李海洋. (2017). 管理学研究中的内生性问题及修正方法. *管理学季刊*, *3*(35), 20–47.

温忠麟. (2017). 实证研究中的因果推理与分析. *心理科学*, *40*(1), 200–208.

温忠麟，陈虹熹，叶宝娟，方杰，蔡保贞. (2022). 新世纪20年国内测验信度研究. *心理科学进展*, *30*(8), 1682–1691.

温忠麟，侯杰泰，马什赫伯特. (2004). 结构方程模型检验：拟合指数与卡方准则. *心理学报*, *36*(2), 186–194.

熊猛，刘若瑾，叶一舵. (2020). 单亲家庭儿童相对剥夺感与心理适应的循环作用关系：一项追踪研究. *心理学报*, *53*(1), 67–80.

张晓敏，李亚男，徐慧. (2022). 教育如何影响流动人口的社会融入?——基于义务教育法实施的自然实验. *教育经济评论*, *7*(2), 20–43.

赵西亮. (2017). 教育、户籍转换与城乡教育收益率差异. *经济研究*, *52*(12), 164–178.

朱琪，陈臻，李博湛，杨洋. (2022). 老龄化背景下人工智能对人工成本的影响：来自中国31个省份的证据. *华南师范大学学报(社会科学版)*, (2), 142–158.

Aliprantis, D., & Richter, F. G.-C. (2020). Evidence of neighborhood effects from moving to opportunity: Lates of neighborhood quality. *Review of Economics and Statistics*, *102*(4), 633–647.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, *38*(1), 37–72.

Bollen, K. A. (2018). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, *54*(1), 31–46.

Burns, R. A., Crisp, D. A., & Burns, R. B. (2019). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the cross-lagged panel and random-intercept cross-lagged panel frameworks. *British Journal of Educational Psychology*, *90*(1), 77–91.

Certo, S. T., Busenbark, J. R., Woo, H., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, *37*(13), 2639–2657.

Cole, D. A., & Maxwell, S. E. (2003). Testing meditational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–577.

Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of Consulting and Clinical Psychology*, *84*(8), 738–751.

Fang, J. Y., Wen, Z. L., & Hau, K.-T. (under review). Appropriate modeling of endogeneity in cross-lagged panel models: A comparison of auxiliary and model implied instrumental variables.

Fang, J. Y., Wen, Z. L., Ouyang, J. Y., Huang, G. M., & Liu, X. Q. (2022). The reciprocal longitudinal relationship between body mass index and subjective well-being in China. *Social Science & Medicine*. *297*, 114829.

Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2020). Latent variable GIMME using model implied instrumental variables (MIIVs). *Psychological Methods*, *25*(2), 227–242.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurements*, *5*(4), 475–592.

Heckman, J. J. (1979). Sample selection bias as a specification error. Applied Econometrics, *47*(1), 153–161.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2020). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, *47*(1), 105–143.

Kazuma, S. (2021). Unhappy and happy obesity: A comparative study on the United States and China. *Journal of Happiness Studies*, *22*(3), 1259–1285.

Lebenbaum, M., Laporte, A., & de Oliveira, C. (2021). The effect of mental health on social capital: An instrumental variable analysis. *Social Science & Medicine*, *272*, 113693.

Lu, H., Tong, P., & Zhu, R. (2019). Longitudinal evidence on social trust and happiness in China: Causal effects and mechanisms. *Journal of Happiness Studies*, *21*(5), 1841–1858.

Maydeu-Olivares, A., Shi, D., & Fairchild, A. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, *25*(2), 243–258.

Patel, P. C., & Cooper, D. (2014). Structural power equality between family and non-family TMT members and the performance of family firms. *Academy of Management Journal*, *57*(6), 1624–1649.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Semadeni, M., Withers, M. C., & Certo, S. T. (2014). The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, *35*(7), 1070–1079.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, *44*(4), 571–585.

Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics*, *79*(2), 348–352.

Streeter, A. J., Lin, N. X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., & Henley, W. E. (2017). Adjusting for unmeasured confounding in nonrandomized longitudinal studies: A methodological review. *Journal of Clinical Epidemiology*, *87*, 23–34.

Tong, E. M. W., Lum, D. J. K., Sasaki, E., & Yu, Z. (2019). Concurrent and temporal relationships between humility and emotional and psychological well-being. *Journal of Happiness Studies*, *20*(5), 1343–1358.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relation. *Psychological Methods, 24*(5), 637–657.

Wiedermann, W., & von Eye, A. (2020). Reciprocal relations in categorical variables. *Psychological Methods*, *25*(6),

708–725.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd Edition). Cambridge, MA: MIT Press.

The endogeneity issue in longitudinal research: Sources and solutions

FANG Junyan[1], WEN Zhonglin[2]

([1] *School of Leisure Sports and Management, Guangzhou Sport University, Guangzhou* 510500, *China*)

([2] *School of Psychology/Center for Studies of Psychological Application, South China Normal University, Guangzhou* 510631, *China*)

Abstract: The endogeneity issue is often induced by limited causes in cross-sectional studies, while multiple sources of endogeneity may exist in longitudinal research. Bivariate longitudinal studies have been widely utilized in the exploration of causation in behavioral and psychological sciences, while the endogeneity issue in longitudinal research has received little attention, which may impair estimation accuracy. In this study, we elaborate on the main sources of endogeneity in longitudinal research, including omitted variables, self-selection, sample selection, measurement error in the predictors, dynamic panel, and reciprocal relation. Taking the Cross-lagged panel model as an example, the empirical analysis reveals the impact of endogeneity. We also discuss the feasibility and rationality of applying instrumental variables in the longitudinal model to deal with the endogeneity. The purposes are to draw more attention to the endogeneity issue and to promote better use of longitudinal models in causal analysis.

Keywords: endogeneity, longitudinal research, cross-lagged panel model, instrumental variables

附录：R 语句

```
library(lavaan)
data<-read.table(file=" 11.DAT",header=F)
colnames(data)<-c("id","x1","x2","x3","y1","y2","y3")
model <- '
etax1 =~ 1*x1; etax2 =~ 1*x2; etax3 =~ 1*x3; etay1 =~ 1*y1; etay2 =~ 1*y2; etay3 =~ 1*y3
etax2 ~ a1*etax1; etax3 ~ a1*etax2; etay2 ~ a2*etay1; etay3 ~ a2*etay2
etay2 ~ c1*etax1; etay3 ~ c1*etax2; etax2 ~ c2*etay1; etax3 ~ c2*etay2
x1 ~~ varxx*x1; x2 ~~ varxx*x2; x3 ~~ varxx*x3
y1 ~~ varyy*y1; y2 ~~ varyy*y2; y3 ~~ varyy*y3
etax2 ~~ varx*etax2; etax3 ~~ varx*etax3;etay2 ~~ vary*etay2
etay3 ~~ vary*etay3;etax1 ~~ varx1*etax1; etay1 ~~ vary1*etay1
etax1 ~~ cov1*etay1;etax2 ~~ cove*etay2; etax3 ~~ cove*etay3
x1 ~ 0*1; x2 ~ 0*1; x3 ~ 0*1; y1 ~ 0*1; y2 ~ 0*1; y3 ~ 0*1
etax1 ~ 1; etax2 ~ 1; etax3 ~ 1; etay1 ~ 1; etay2 ~ 1; etay3 ~ 1
etax2 ~ zx*x1;etax2 ~ zx*y1;etay2 ~ zy*x1;etay2 ~ zy*y1
x1 ~~ y1; etax2 ~~ etax3;etax2 ~~ etay3;etay2 ~~ etax3;etay2 ~~ etay3'
fit <- sem(model, data = data)
summary(fit, standardized = TRUE)
```